MASSACHUSETTS INSTITUTE OF TECHNOLOGY
LINCOLN LABORATORY

ARTIFICIAL NEURAL NETWORKS FOR SEISMIC DATA INTERPRETATION

SEMIANNUAL TECHNICAL SUMMARY

1 MAY – 30 NOVEMBER 1990

ISSUED 17 MAY 1991

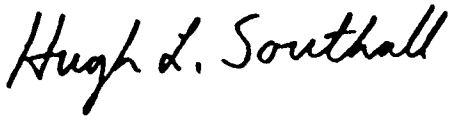
LEXINGTON MASSACHUSETTS

This report is based on studies performed at Lincoln Laboratory, a center for research operated by Massachusetts Institute of Technology. The work was sponsored by the Department of the Air Force under Contract F19628-90-C-0002.

The ESD Public Affairs Office has reviewed this report, and it is releasable to the National Technical Information Service, where it will be available to the general public, including foreign nationals.

This technical report has been reviewed and is approved for publication.

FOR THE COMMANDER



Hugh L. Southall, Lt. Col., USAF
Chief, ESD Lincoln Laboratory Project Office

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
LINCOLN LABORATORY

ARTIFICIAL NEURAL NETWORKS FOR SEISMIC DATA INTERPRETATION

SEMIANNUAL TECHNICAL SUMMARY

1 MAY – 30 NOVEMBER 1990

ISSUED 17 MAY 1991

LEXINGTON MASSACHUSETTS

EXECUTIVE SUMMARY

This is the first Semiannual Technical Summary Report of the MIT Lincoln Laboratory Artificial Neural Networks for Seismic Data Interpretation project.

Introduction

Seismic surveillance applications were reviewed and data interpretation tasks were selected for initial neural network experimentation. The selected tasks are estimation of signal arrival time (time picking), labeling of seismic phases, and recognition of typical and atypical events on a regional basis. Basic seismology and surveillance techniques are reviewed in this report and preliminary experimental results are summarized.

Data Base

We are using two types of data. Seismic waveform data with associated parametric information are being provided by SAIC in San Diego, CA. Parametric data for a much larger data set are being obtained by remote access to an on-line data base at the Center for Seismic Studies (CSS) in Arlington, VA. All the data are from NORESS and ARCESS arrays in Scandinavia and were processed by the IMS (Intelligent Monitoring System) regional seismic surveillance system. At the start of the contract SAIC provided an initial waveform data set for exploratory experimentation. While using it, our waveform data requirements and formats were worked out with SAIC and the first installment of waveforms for 50 events has now been received.

Arrival Time Estimation

Arrival time estimation experiments concentrated on using perceptrons for arrival time estimation. The initial waveform data set was used and we concentrated on the *Pn* phase. Differences between automatic and human picks were found to be small (a fraction of a second). Simple single-layer perceptrons worked as well as more complicated topologies, and a few simple signal features seemed to capture all of the relevant information for time picking. The networks did improve on other automatic picks for this data base, but the improvements were small and probably not of operational significance.

A preliminary review of IMS statistics in the on-line CSS data showed that they are very different from those of our initial data set. They range from a few to several seconds on average, depending on phase type. This makes it difficult to draw hard conclusions from our initial experiments. For real IMS data it does appear possible to make operationally significant improvements.

Phase Identification

Phase identification experiments concentrated on sonograms from vertical seismometers and three-component autoregressive signal representations for phase identification. ART-2 neural networks and unsupervised learning were used for the sonogram experiments. Radial basis function (RBF) network and Gaussian classifiers were used in the three-component experiments.

The sonogram experiments used eight NORESS events from our initial waveform data set, each event with four reported phases. Several sonogram preprocessing options and different numbers of ART-2 categories (controlled by the vigilance parameter) were tried. Correct recognition percentages up to 84–88 percent were obtained for *Pn*, *Pg*, *Sn*, and *Lg* using 8–10 ART-2 categories. Gaussian classifier experiments using autoregressive signal representations were also done with 10 events from the initial waveform data base; a success rate of about 90 percent was achieved.

Gaussian classifier and RBF neural network experiments using autoregressive signal models were performed with 152 ARCESS phase arrivals contained in the 50-event data base recently received from SAIC. Phase categories for these experiments included *Pn*, *Pg*, *Px*, *Sn*, *Sx*, and *Lg*. The success rates were in the 50–60 percent range for the Gaussian classifier. RBF success rates were higher on training data, especially for high-complexity RBFs, but the RBF success rates on separate test data were not as good as the Gaussian classifier success rates.

Future experiments will use more signals, investigate additional preprocessing options, and include parameters (features) that are produced by the IMS system in addition to sonograms or autogregressive models. The goal is to improve performance to a level useful in the IMS.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

1. INTRODUCTION

Networks and arrays of seismometers can be used to detect and locate seismic events and to distinguish between different types of events. Figure 1 is a block diagram of a system to process seismic signals and to perform these functions. It is patterned after the Intelligent Monitoring

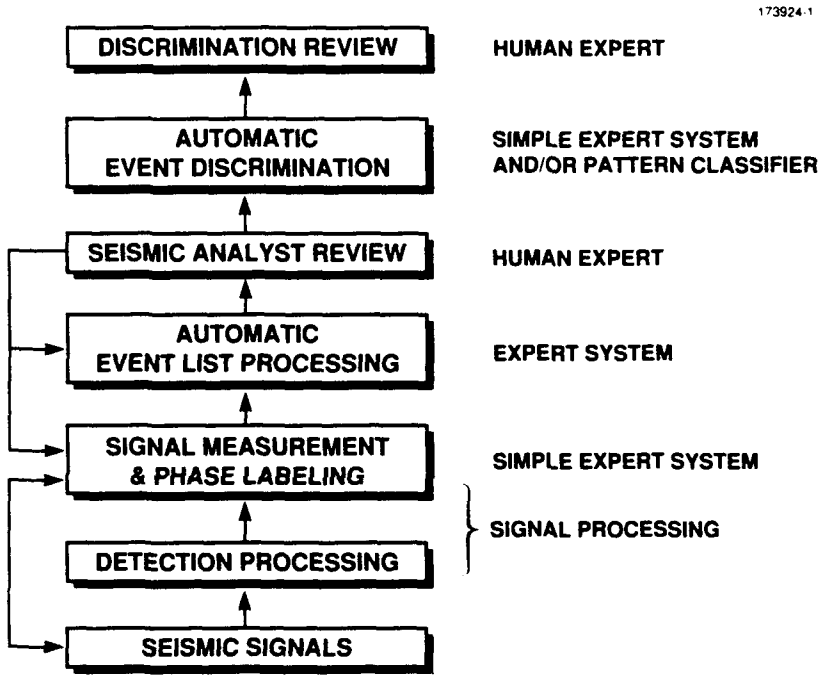


Figure 1. Functional elements in a modern seismic surveillance system.

System (IMS) [1,2] developed by DARPA for interpreting seismic data. That system, which evolved from an earlier system [3] that did not exploit expert system techniques, is designed to detect and locate seismic sources, determine their magnitudes, and provide information to help identify them as earthquakes, chemical explosions, or underground nuclear explosions. The objective of the Lincoln Laboratory Artificial Neural Networks for Seismic Data Interpretation program is to develop neural-network components for possible integration into such a system.

1.1 Seismological Background

Because seismic data interpretation for nuclear test monitoring is a specialized application area, we are including in this First Semiannual Technical Summary Report a brief seismological tutorial to introduce terminology and to put our work into context.

Seismic signals from large events can be detected at very large distances (tens of thousands of kilometers). But small events, which are important for underground nuclear test monitoring at low yields, are detectable only at relatively small distances. Therefore, recent work on nuclear test surveillance has concentrated on using seismometers located within regional distances (typically less than 1000–2000 km) of the area being monitored. For this reason, our work will emphasize the interpretation of regional data.

Seismometer arrays are used to estimate signal propagation speed and direction of arrival and to increase signal-to-noise ratios (SNRs) by phased-array methods. In current practice for regional arrays, most array seismometers are vertical instruments that respond only to the vertical component of particle motion. However, the arrays may also contain a few three-component seismometers, each consisting of two horizonal and one vertical seismometer at the same location. The three component sensors provide additional polarization and directional information. Small "regional" arrays [4] with both vertical and three-component seismometers, designed especially for use at regional distances, have now been installed in Scandinavia (NORESS, ARCESS) and Germany (GERESS) and will be primary sources of data for our neural-network research.

Transient seismic sources produce several wave packets at regional distances, called phases. The theoretical possibilities are *P, S, L,* and *R* phases, corresponding to compressional body waves, shear body waves, and two different surface wave types (Love and Rayleigh). There are many variations of type and nomenclature. *Pn, Pg, Sn, Sg, Lg,* and *Rg* are used to denote *P, S, L,* and *R* wave types that are observed at regional source-receiver distances. The "*n*" and "*g*" subscripts denote different propagation paths. The *g* phases propagate entirely within the crust of the earth. The *n* phases travel deeper, arrive sooner for most distances of interest, and propagate along the crust mantle boundary for much of their path length. One caveat to be noted is that although *Lg* implies a Love wave travelling through the crust, this im[illegible]tant phase is now believed to included more than Love waves. The different phases are roughly distinguishable on the basis of propagation speed, relative arrival times, polarization characteristics, and frequency content, but algorithms for phase identification are by no means perfect.

Different propagation paths greatly influence the appearance of seismograms and cause significant variability in the recorded waveforms. The most obvious effects are the relative prominence of different phases in the seismograms from different directions or distances, but more subtle differences (e.g., frequency content) are also important. Propagation-induced effects greatly complicate the interpretation of seismic signals, since they can often obscure or be confused with source effects. Propagation-induced variability is particularly strong for observations at regional distances. It is important to be alert to this while developing neural networks to aid in seismic data interpretation.

Despite these complications, seismic event location is a routine process. Locations are estimated by fitting predicted phase arrival times to measured arrival time. For example, the P-to-S time from a single seismogram at regional distances provides a good estimate of the distance to the source. Arrival time measurements from several phases and from several receiver sites can be combined to estimate the event location, including depth. Direction information and wave speed estimates from arrays and three-component seismometers are also used for event location.

Seismic sources are located in three dimensions: latitude, longitude, and depth. This location is the event hypocenter. The epicenter is the geographic location consisting of only the latitude and longitude. Although event location is routine, improvements in accuracy are always being sought, especially for depth, which is critical for event identification and is usually not accurately determinable for that purpose.

After an event is located, seismic magnitudes (or other measures of size) are estimated. This involves signal amplitude measurements and empirical corrections to remove propagation losses. There are several different magnitude scales, and more than one magnitude is usually estimated. The relative size of an event on the different magnitude scales can be diagnostic of the event type, and for nuclear tests it can be used to estimate yield. For nuclear test monitoring systems, the final data interpretation step is classification of the event as a nuclear explosion, chemical explosion, earthquake, or unknown source.

Figure 1 shows this flow of processing and the nature of the processing at each stage. The general flow consists of: signal detection processing; seismic phase labeling and parameter measurement; event processing to determine the time, location, and magnitude of seismic events; and event classification.

Much of the processing outlined in the figure is automated, but human seismic analysts play an important role in the system. They review automatic processing results and make corrections. The corrections include changing signal onset time estimates, changing assigned phase identifiers, or changing which detections are grouped together and associated with a single seismic source (event). This is a complex process in which the analysts bring many different kinds of knowledge to bear, including their ability to recognize events from looking at the seismic waveforms. Examples of this are recognizing that an event is an explosion from a specific quarry or nuclear test site or knowing that natural earthquakes from a specific seismic region almost always look similar. This recognition is often receiver- as well as source-location specific and is hard to quantify.

The last stage of processing is event identification (discrimination). It is generally based on a relatively small number of criteria which have underlying intuitive or physical justifications and have evolved empirically. Emphasis has been on event parameters (magnitude, depth, epicenter, etc.) and on the distribution of signal energy in frequency and into different seismic phases.

1.2 Research Plan

Figure 2 indicates one way that we envision an artificial neural network's being integrated into an overall surveillance system. The basic idea is to develop neural networks that will perform specialized functions that can be integrated into the present IMS as it evolves.

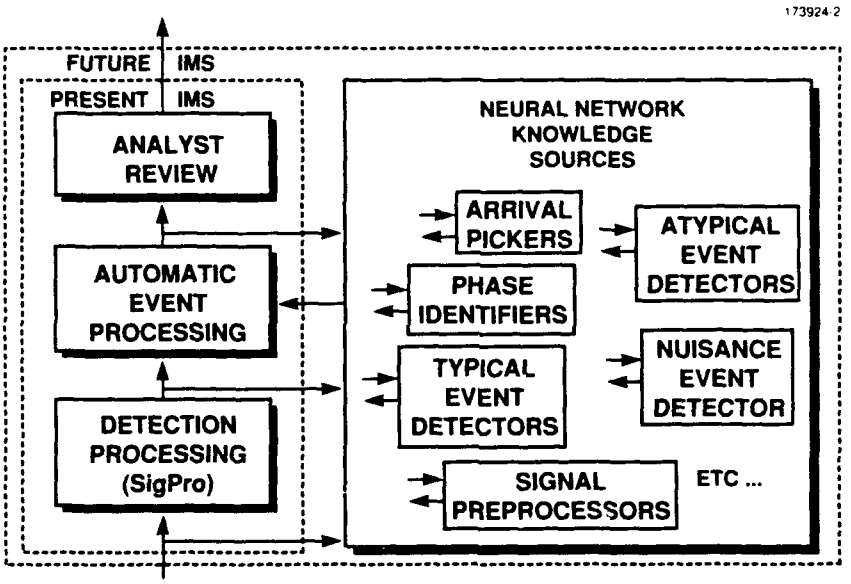


Figure 2. Integration of neural networks into a seismic surveillance system.

We have decided to concentrate on relatively low-level data interpretation tasks rather than the ultimate discrimination task, which is deciding whether an event is an earthquake, chemical explosion, or nuclear explosion and, if nuclear, what the yield is. Thus, the discrimination and discrimination review blocks of Figure 1 are omitted from Figure 2. Also, in Figure 2 the signal measurement and phase-labeling functions are considered to be part of detection processing. All of these are accomplished in the IMS by a "SigPro" software package.

The decision to concentrate on tasks other than the top-level discrimination task was made for several reasons. One is that classification of an event as a nuclear explosion is a politically sensitive decision. It must be explained in human terms and the explanation should be based on physical theory. Although this does not exclude the use of neural networks, we felt that we could identify better uses for networks in the seismic application. Another potential difficulty with the discrimination problem is the relatively small number of representative events, especially the critically important nuclear explosions. This is exacerbated by seismic propagation effects that are highly variable and difficult to eliminate; there could be serious difficulties with generalizations made on the basis of test data. We can better predict the performance of discriminants with a physical basis under new circumstances than we can a discriminant that is almost entirely statistical, and neural-network discriminants are essentially statistical.

Figure 2 cites several functions that we have identified as candidates for neural-network implementation and on which our research is concentrating: arrival-time picking, phase identification, typical- and atypical-event detection, and nuisance-event detection. With the exception of the last, these all share the important characteristic that occasional errors by the network will be routinely detected by human analysts and will cause no significant political or operational problem. The emphasis is on improving the automatic processing in the system so that overall performance may be improved and the analyst load reduced. However, the system is still designed so that occasional errors have minimal impact. The functioning of the neural networks need not be "explained" (require a theoretical basis) any more than the traditional event detectors do; it is only required that performance be statistically satisfactory.

Our initial research will concentrate on arrival-time picking and phase identification, for several reasons. Discussions with seismic experts in Norway and the United States who are familiar with operational systems confirmed that arrival time estimates and phase identifications made by automatic algorithms must often be corrected by human analysts. Thus, improvements in the automatic performance of these functions is desirable to reduce the analyst workload and make the overall system more automated.

An advantage of time-picking and phase-identification applications is that minimal requirements are placed on source type (explosion, earthquake, etc.) for training or testing. By definition, a phase has to do with propagation. Thus, the event type should (must) not matter very much for phase identification. Similarly, time picking may depend on signal and noise characteristics but not explicitly on the type of event. Because of this, it will be relatively easy to accumulate training and test data sets.

The neural-network time-picking and phase-identification modules will be developed to improve on the picks and identifications produced by existing IMS algorithms. We will depend on SigPro to detect the events and provide initial guesses of these parameters. For phase identification we will explore features that are routinely calculated by the IMS and additional features that we will extract directly from seismograms. Initial work on phase identification will concentrate on

"out-of-context" identification; that is, each phase arrival will be treated locally without regard for decisions made about other arrivals before or after the one under consideration.

The IMS in Norway is routinely used to process data from NORESS and ARCESS. Seismic analysts routinely review and correct the output from the automatic IMS. All changes and corrections made by the analysts are recorded. Thus, we will have available not only time picks and phase identifications made by the automatic system but also the corrections made by the analyst. For our neural-network development work we will consider the analysts' results to be correct. From one viewpoint this can be open to question, since many of the decisions are fuzzy and different analysts may not agree, especially for events with low SNR or phases in the coda of earlier phases. (The coda of a phase is lower-level energy that continues to arrive after the first burst and is usually attributed to scattering and multipath propagation effects.) However, the analyst's decision is the best available view of fact. Moreover, as long as analysts are in the loop and can modify the output of the automatic system, it is clear that the task of the automatic system is to minimize the number of changes the analyst will make. In practice the analyst's decisions must be treated as correct.

Once we have completed our exploratory development of time-picking and phase-identification networks, we plan to expand the work to include "in-context" phase identification and typical and atypical event recognition. These have in common the fact that the entire seismogram, not just an isolated phase arrival, will be taken into consideration for classification purposes.

Automatic recognition of atypical and typical events is not now part of the IMS. (There is a script matching component intended to classify events on the basis of signal-to-noise measurements on a set of standard array beams plus a few additional parameters, but this is a very experimental element of the system that is not integrated into routine operations [1].) However, if neural networks could learn to recognize atypical and typical events, the system might be modified to use this capability. At the least, the analyst might be alerted to pay attention to specific events. The idea is simple and related to the fact that human analysts often learn by experience to identify events simply by looking at their seismograms. For example, for a given receiver site events from a particular seismic region may have a very characteristic appearance. This is both source-region and receiver specific. It is also true that not all events from a region will match the pattern. Our plan is to investigate how networks might learn to recognize a typical event and, perhaps more important, recognize an atypical event. For example, if an explosion took place in an active earthquake area where there had been no known previous explosions, a network could alert the analyst that the signals from this event seemed strange; this focus of attention might be quite helpful.

Nuisance-event recognition is a specialized atypical/typical event-recognition problem that we may investigate. A nuisance event is a real signal, detected by an array, from a very small event very close to the receiver. It is not large enough to be detected at other arrays or to be of any interest. It may be due to local cultural activities, nearby ice fractures, etc. There may be many of these events, and they can consume processor and analyst time. Thus, recognizing them early in the processing chain and eliminating them from further consideration would be useful.

The choice of problems for which we plan to seek neural-network solutions was made during the first few months of this research effort, when we reviewed data sources, discussed needs with potential users, and evaluated the match between seismic problems and neural-network technology.

2. DATA BASES

During this reporting period we have worked with Science Applications International Corporation (SAIC), which is under DARPA contract to provide data for this project, to define seismic signal data bases for our exploratory research, to specify data formats and exchange mechanisms, and to obtain and start using seismic waveform data.

This was done in two stages. At the very start of the project we were provided with an initial waveform data set that was prepared by SAIC for another purpose; this was the only waveform data available to us for most of this reporting period. While using this data for exploratory experimentation, we defined additional data sets to be created and provided specifically for this project. These newly defined data sets are now becoming available. All data sets consist of data from regional seismic arrays in Scandinavia. We have also started to access and use the IMS parametric data that is on line at the Center for Seismic Studies in Arlington, VA.

During this period we also procured, installed, and started to use Sun SPARC stations as our basic data-analysis and processing tools. These standard workstations are compatible with the DARPA IMS and are also used by the SAIC group that is providing us with seismic data. This commonality assures smooth interactions with SAIC and will facilitate possible transition of our results to the IMS for more extensive evaluation.

2.1 Initial Waveform Data Set

The initial waveform data set provided by SAIC contained a single three-component waveform segment from the NORESS array for each of 73 regional events. Location and magnitude estimates were provided for each event. The 73 events included: (1) 23 earthquakes, (2) 39 chemical explosions (20 from the Blasjo mine site and 19 from the Titania mine site in southwest Norway), and (3) 11 events off the southwestern coast of Norway. All were between 300 and 500 km from NORESS. This data base was constructed by SAIC under contract to HNC, Inc. of San Diego, CA, for a neural-network research effort to distinguish between chemical explosions and earthquakes. It was provided to us with HNC concurrence. The data were provided in Seismic Analysis Code (SAC) format to allow us to evaluate that option for waveform data exchange and to provide us an initial data set for initial experimentation.

SAC is an interactive time-series-analysis and signal-display software package that was originally developed by Joseph Tull at Lawrence Livermore National Laboratory. It has subsequently been adopted and used for seismic data analysis by several organizations including the SAIC group providing the seismic data for this project. SAC format is the seismic waveform storage format understood by the SAC software. We obtained the SAC code from SAIC along with the initial data. As expected, we have found this format satisfactory; it is already in routine use and allowed us to start work with minimal software effort. Results of our preliminary experimentation with this data set are presented in Chapters 3 and 4.

This initial data base only allowed us to begin exploratory research. It has very limited distance and azimuth coverage, contains data from only one seismic receiver site (NORESS), includes only one three-component sensor from that site, and does not include any of the detecting beams. Because the events in this data set predate the time when IMS came into routine use, it also contains inadequate ancillary parametric data. Specifically, our research requires that we know the automatic time picks and phase assignments made by IMS and how they were changed by human analysts. Thus, while using this initial data set for some preliminary experiments, we also worked with SAIC to define and obtain a data base better tuned to the needs of this project.

2.2 Waveform Data Bases

Working with SAIC, we have defined eight waveform data bases that they will provide to Lincoln Laboratory. Data will be provided from both NORESS and ARCESS whenever both sites appear to have detected the same event. For each event in each data base we will obtain all unprocessed signals and all detecting beams. In each case a seven-minute data segment will be provided, starting 30 seconds before the first phase arrival for the event. Complete parametric data, including all parameters generated by automatic processing and all changes made by analysts, will be provided. The data exchange media will be read/write optical disks.

In addition to waveforms and parametric data in electronic form, SAIC is providing hard-copy information for each event. This includes listings of the parameters, maps showing the geographic location of the event, a copy of the standard display provided to the analyst, and plots of the detecting beams and of three component beams for each phase associated with the event.

The eight waveform data bases have been defined and sized to provide a broad sampling of seismic waveform phenomena and source coverage and to expedite obtaining data that we think will be particularly useful. Once we have obtained and experimented with these initial data sets, we will work with SAIC to determine which ones need to be expanded for more complete training and testing.

The eight data bases and the number of events requested are:

1. High-quality regional analyst-corrected events (50)
2. High-quality regional analyst-accepted events (50)
3. Random selection of analyst-corrected events (50)
4. Random selection of analyst-accepted events (50)
5. Random selection of analyst-rejected events (25)
6. Non-event detections (40)
7. Teleseisms (25)
8. Unusual events (10)

The "high-quality" requirement for the first two data bases means that the request is restricted to events for which the waveforms have been saved on line by the IMS; this procedure expedites obtaining the data. The first of the eight waveform data bases has been received and we are beginning to use it. Now that all formats and necessary software are available, we expect that the rest of this waveform data will be available shortly.

2.3 Parametric Data Bases

In addition to the parameters provided with waveform bases, it has become clear that we should use the much larger parametric data base that is routinely generated by IMS and maintained at the Center for Seismic Studies (CSS) in Arlington, VA. This data base contains lists of all detections made by the NORESS and ARCESS arrays, lists of automatically extracted parameters for each detection, automatically generated event lists, documentation of corrections made by analysts, and revised event lists reflecting the corrections. We need these data to obtain analyst time-correction statistics, phase-identification change statistics, and to experiment with phase identification using parameters routinely generated by the IMS. By using the complete parametric data base we can derive statistics and perform phase-identification experiments with a much larger data set than would be possible if we restricted ourselves to events for which we obtain waveform data.

The original plan was to ask SAIC, which is providing us with waveform data, to provide additional parametric data. However, the parametric data are now on line at CSS and accessible to us by computer network. This is a more convenient and flexible option and is the mode we are now using.

Chapter 3 includes arrival-time estimation statistics that were compiled using this data base.

3. ARRIVAL-TIME ESTIMATION

Arrival-time estimation involves determining, as accurately as possible, the onset times of seismic phases. It is important because arrival-time estimates are used to obtain event locations. Arrival-time estimation can be difficult if the SNR is poor or if the newly arriving phase has a smaller amplitude than the preceding phase.

Algorithms for seismic phase detection typically begin by computing a simple energy measure, such as the ratio of a short-term average to a long-term average (STA/LTA) of the energy in the seismogram. When this ratio exceeds a predetermined threshold, a detection is declared. The threshold-crossing time may then be used as an arrival-time estimate, or the arrival time may be further refined by various algorithms. In the IMS [1], the time pick is refined using a technique [3] that involves analyzing the peaks and valleys of the signal near the threshold-crossing time. An analyst then reviews the time picks and corrects them as necessary.

Our goal is to develop neural networks to improve the time estimates generated by the IMS automatic algorithms and to reduce the number of corrections the analyst must make. The inputs to our neural network could include the signal itself (perhaps filtered, scaled, and rectified), the initial time estimate generated by the IMS and various parameters generated by the signal-processing software (e.g., rectilinearity, dominant frequency, SNR). The output of the neural network could be a new time estimate or an indicator as to the expected accuracy of the automatic time pick.

Our work on this project so far has concentrated on better understanding the nature of the corrections made by the analyst and determining the types of inputs that produce the most accurate time picks from the neural network. To better understand the nature of the analysts' corrections, we have begun collecting statistics on the size of the corrections for different phases. These statistics are given in Section 3.1. To determine the best types of inputs to give the neural network, we have performed some experiments with the initial data set described in Section 2.1. The results of these experiments are summarized in Section 3.2.

3.1 Analyst Timing Corrections

Since our primary goal is to use a neural network to reduce the number of arrival-time corrections the analyst must make, it makes sense to study the corrections currently being made. We have learned to access the CSS data base [5] using the SQL language and the SQL*Plus tool [6], and have used this knowledge to extract on-line statistics concerning analyst timing corrections. In this section, we describe the statistics we have obtained.

The statistics given in this section were extracted from the CSS data base IMS1, which contains parameter information about a large number of seismic events detected at the NORESS and ARCESS seismic arrays. While it would be possible to collect statistics on all the detections in the data base, this may not be meaningful; some of the detections were never associated with any particular event or any other phases, thus it is unlikely that the analysts made any effort to

correct them. Therefore, when collecting statistics on arrival-time corrections, we have imposed the following restrictions:

- The phase arrived sometime after January 1, 1990. At the time of running this experiment, the data base contained arrivals through October 1990, so these statistics include about 10 months of data.
- The phase must be associated with some event (and therefore with at least one other phase); this association must be established or confirmed by the analyst.
- The phase must be detected by the automatic system and kept by the analyst. Phases added by the analyst or discarded by the analyst are not included.
- The phase must be identified as *Pn*, *Pg*, *Sn*, or *Lg*; the identification must be established or confirmed by the analyst.

All signals satisfying the above restrictions are included in the statistics discussed below. We have not yet included SNR or other quality estimates in these statistics, although collecting additional statistics as a function of signal quality might be useful in the future.

Table 1 shows the statistics we have collected as a function of seismic phase.

TABLE 1

Statistics on Analyst Timing Corrections

Phase	Number of Signals	Percent Changed	RMS Change (s)
Pn	5750	42	2.65
Pg	1020	31	1.92
Sn	1345	49	16.57
Lg	5810	52	5.68

These statistics indicate that the corrections made by the analyst are significant, and suggest that a system for reducing the number of corrections would significantly reduce the work of the analyst. Also, since the size of the corrections varies significantly from phase to phase, a different network structure or a different parameter set may be necessary for different phases. In addition, the *Pn* and *Lg* phases are detected far more often than *Pg* and *Sn*, so it makes sense to concentrate on these phases in our early work.

3.2 Initial Experiments

Ideally, we would like to perform experiments with seismograms obtained from the IMS data base discussed in Section 2.2. Unfortunately this was not possible since we have only recently begun to receive these data. Therefore, the results in this section are based on events in our initial data set (Section 2.1). As will be seen, for this data set the accuracy of automatic time picking is better than it appears to be on average for the IMS data. This limits our ability to determine how well the neural network will perform in combination with the IMS, but nevertheless allows us to experiment with different types of inputs to the neural network and see which inputs give us the most accurate time estimates.

The approach in our initial experiments has been to first compute a crude estimate of the arrival time using an STA/LTA detector, then use the neural network to refine this estimate. Our longer-term plans are to use the estimate generated by the IMS rather than our own STA/LTA estimate. In our initial experiments, we have used data from the 1-s interval preceding the STA/LTA detection. This interval contained the true *Pn* arrival time for most of the seismograms. However, based on the statistics we now have concerning analyst corrections in the IMS,

we will change the length of this window when we begin using the IMS data. Our initial experiments concentrated on the *Pn* phase, since it is usually the first to arrive and is often easiest to detect with an STA/LTA detector. We plan to consider other phases in the future, probably starting with *Lg*.

The neural network structure in our initial experiments is a multilayer perceptron trained with back-propagation [7]. The network inputs include the data itself, filtered and rectified, various energy measures, and other parameters such as rectilinearity. The inputs are described in more detail below. The network output is a single analog output that indicates the position of the new time estimate within the chosen data window. We have chosen to use a single analog output (rather than a series of binary outputs) because the back-propagation algorithm minimizes the mean-square difference between the actual and desired outputs. With a single analog output indicating the onset time, the back-propagation algorithm will minimize the RMS estimation error.

While we may try other types of networks in the future, we have started with the perceptron because it is a classic architecture that has been successfully applied to a large variety of problems. In the experiments performed to date, we have tried 1-, 2-, and 3-layer perceptrons with comparable results, suggesting that the 1-layer perceptron is adequate for the data we have used so far; the results given in Section 3.2.1 were obtained using a 1-layer network. Our effort to date has concentrated not on the network structure itself, but on choosing signal features to maximize the network performance. Feature options and multilayer networks must be reinvestigated for IMS data.

3.2.1 Input Representation

This section contains results obtained with perceptrons and several different types of signal feature vectors. The raw data were signals from 52 *Pn* phases with good SNR. The "true" arrival times were picked by an experienced seismic analyst. The experiments used only one vertical seismometer (except where rectilinearity is included, in which case a three-component seismometer was used). The numerical values for the RMS errors may not be realistic performance estimates by themselves, but should be useful in comparing different approaches.

It is usually important to use different sets of data for network training and testing so that the testing results are indicative of performance that will be obtained on new data. The procedure we have used is the "leave-one-out" method, designed to make maximum use of a small data set. In this procedure, we first train the neural network using 51 of the 52 signals, then test it on the one signal that was not used for training. Then we repeat this procedure a total of 52 times, each time leaving out a different signal. Finally, we compute the RMS time error, averaged over all 52 trials.

Table 2 shows a list of different inputs we have used and the RMS time error resulting from these inputs. For comparison purposes, the RMS error resulting from using the STA/LTA estimate directly is 0.41 s; the RMS error from using the STA/LTA estimate but correcting for its bias (i.e., by subtracting the mean time difference between the actual arrival time and the STA/LTA threshold crossing) is 0.22 s.

TABLE 2

Effect of Different Inputs on Time Picking

Inputs	RMS Time Error (s)
1. Signal (40 points)	.16
2. Envelope (40 points)	.13
3. List of Peaks (10 peaks)	.17
4. Energy in 0.5-s window	.15
5. Energy in 0.5-s envelope	.12
6. Maximum SNR	.14
7. Rectilinearity (40 points)	.18
8. 5 and 6 above	.11

The first item in the table, labeled "signal", uses data from the 1-s interval preceding the STA/LTA threshold crossing. The data is filtered to 8–16 Hz and rectified. The item labeled "envelope" uses the envelope of this filtered, rectified signal—that is, the peaks of the signal are preserved and samples in the intervals between the peaks are replaced with a linear interpolation between the adjacent peaks. This forces the neural network to concentrate more on the overall energy structure of the signal and less on individual peaks and valleys. The item labeled "list of peaks" uses a list of pairs (time, amplitude), each representing a peak in the signal rather than a value at each sample time. "Rectilinearity" listed in this table is computed as a function of time, using a sliding 2-s window, and therefore varies very slowly over the data interval in question By comparing the results of these experiments, note that the envelope gave better performance than the raw data, the list of peaks, or the rectilinearity.

Equally good performance can be obtained by computing a very simple energy sum over a specified interval, thus having only one or two inputs to the neural network rather than 40. The results listed in the table are for energy sums over only 0.5 s rather than 1 s; we have tried intervals in the range of 0.25 s to 2.0 s and found that 0.5 s gives the best results.

These results suggest that the essential information the network is extracting from the signal is the amount of energy immediately preceding the STA/LTA threshold crossing. If this energy is low, the arrival time is near the end of the interval (near the STA/LTA crossing). If this energy is large, the arrival time is near the beginning of the interval, further from the STA/LTA crossing.

Theoretically, better results might be obtained by using several inputs simultaneously rather than individually. In fact, our experiments suggest that combining the inputs does not help significantly. The result is usually close to that obtained with the best of the individual inputs, and is often worse. The best result we have obtained by combining inputs uses the sum of the energy in the envelope and the maximum SNR (over the entire *Pn* phase, not just in a specified interval); this result is 0.11 s, only slightly better than using the energy sum alone. Nevertheless, this is a factor of 2 better than using the STA/LTA detector alone, even with the bias correction.

We conclude from these experiments that for this data base the parameters summarizing the energy content of the signal contain the essential information needed to correct the arrival-time estimate, and that using these parameters in a neural network gives better results than using the STA/LTA alone. However, because the IMS analyst correction statistics are so different from those in this initial experiment, it is difficult to predict the best network structure and signal features to use for the IMS data and how much improvement neural networks will provide. Now that IMS data is available to us, these questions are being addressed.

4. PHASE IDENTIFICATION

Seismic phase identification is the problem of labeling the phase (e.g., *Pn, Pg, Sn, Lg, T,* or *N*) of a waveform, given the waveform and an indication of when the phase arrives (which was the focus of Chapter 3). Phase identification can proceed directly from the raw waveforms recorded at the seismic instruments, from waveforms which have been preprocessed in some manner, or from waveform features that are automatically produced by the IMS before analyst intervention. We plan to experiment with all of these alternatives.

Phase identification can be done in or out of context. For example, human analysts use contextual information to identify phases, that is, they consider the ordering and arrival times of surrounding phases to devise a meaningful interpretation consistent with all the information. At the other extreme are some of the front-end algorithms in the IMS that classify each arrival with no context at all. Some expert system elements of the IMS also use context in making phase identifications. We are working on both in and out-of-context phase classification. An advantage of context-free classification is that it may be more robust when dealing with multiple events in which phase arrivals are interleaved. Our overall goal is to improve on the performance of the automatic phase-classification portion of the IMS and to reduce the number of changes that must be made by the human analyst.

4.1 Data Representation

This section summarizes two signal-representation approaches that we are investigating for phase labeling: a sonogram-based approach that has been applied to single-component vertical waveforms, and an autoregressive method that has been applied to three-component waveforms. Eventually we may choose one of these representations over the other, or we may find it appropriate to keep both of them.

We also include in this section a brief discussion of our planned use of the signal parameters routinely generated by IMS.

4.1.1 Sonograms

Techniques for machine learning and recognition can be borrowed from neural-network image-processing paradigms if the waveform data is transformed into a two-dimensional image-like representation. One such transformation converts a one-dimensional single-component seismic trace to a two-dimensional sonogram (spectral energy vs. time). Other investigators [8,9,10] have considered similar approaches. In some cases, the other investigators first gained experience recognizing signal types in the sonograms, then constructed templates for each of the categories they were interested in matching. Subject to the development of appropriate similarity measures, new events are then compared to the templates and automatically classified. Our approach is different in that neural networks will be used to generate the templates, thus reducing the need for an independent expert.

Automating this process can also rapidly indicate when a particular data representation naturally separates the phases into distinct classes. We are now using this approach to perform phase identification with individual vertical seismograms. Future alternatives include array-average sonograms, beam sonograms, and radial/transverse sonograms. We have concentrated on the vertical component because such data is widely available and positive results would be widely applicable.

4.1.2 Autoregressive Modeling

A natural way to investigate the additional information available from three-component stations is through the covariance structure of the channels. Correlation matrices for different frequency bands are the starting point for the polarization analysis [11] that is included in the IMS. Another approach to representing the polarization and frequency information in the three-component seismometers is autoregressive (AR) modeling, which we are investigating. The goal of AR modeling is to find a small set of parameters which preserves the information content of the waveforms, but uses many fewer parameters than the number of data points. These parameters can then be used for phase identification.

In AR modeling, we assume that the recorded signal, s_n, is a linear combination of past values and some input u_n;

$$s_n = -\sum_{k=1}^{p} a_k s_{n-k} + G u_n \quad , \tag{1}$$

where G is a gain factor. In this single-dimensional case, there are p degrees of freedom since the filter is of order p. The input, u_n, can be either the unit impulse in a deterministic system or white noise in a stochastic environment (since they have the same autocorrelation and the same spectrum). Minimizing the error defined as

$$E = \sum_n (s_n + \sum_{k=1}^{p} a_k s_{n-k})^2 \tag{2}$$

yields a set of p equations with p unknowns, namely the filter coefficients $a_k, 1 \leq k \leq p$. The filter coefficients can then be processed as a feature vector input to a neural-network classifier.

AR modeling is easily generalized to the multidimensional case of a three-component sensor. In this case, $\underline{s}_n$ is now a three-component vector, and the AR model is

$$\underline{s}_n = -\sum_{k=1}^{p} \underline{\underline{A}}_k \underline{s}_{n-k} + \underline{G} u_n \quad , \tag{3}$$

where $\underline{\underline{G}}$ is the noise covariance matrix and the 3×3 $\underline{\underline{A}}$'s are the matrix parameters defining the filter. Now there are $9p$ equations and unknowns in addition to an assumed model for the noise. Although this is a great reduction in order from the number of waveform samples, it may still result in a large number (e.g., 50-100) of parameters.

One excellent way to reduce the dimensionality of the parametric representation is to perform a Principle Components Analysis (PCA, also known as a Karhunen-Loève Expansion). An interesting suggestion, first made by Fukunaga and Koontz [12], is to apply the PCA to the mixture of covariance matrices of the two classes to be discriminated rather than to each covariance matrix separately. This yielded a transform which emphasizes the differences among the classes in addition to reducing the dimensionality of the representation. This approach was used for some of our AR experiments.

4.1.3 IMS Parametric Data

A wealth of useful information for seismic phase identification is generated routinely by IMS. It reports information such as the frequency, SNR, amplitude, rectilinearity, planarity, observed azimuth, and emergence angle of the signals. These parameters are used for phase identification by algorithms and expert systems in the IMS. We plan to experiment with using them for identification by themselves or in combination with sonogram or AR representations. Using the parameters by themselves, we will investigate whether networks can be developed to improve on the IMS out-of-context phase identification algorithms that use the same data. Using the parameters in conjunction with other representation, we hope to determine if important useful information has been discarded by the automatic IMS.

4.2 Classification Methodology

Thus far, we have experimented with Gaussian classifiers from signal processing and Radial Basis Function (RBF) and Adaptive Resonance Theory (ART) classifiers from the field of neural networks.

4.2.1 Unsupervised Neural Networks (ART Networks)

ART1 and ART2 are binary and analog (respectively) classification mechanisms developed by Gail Carpenter and Stephen Grossberg [13,14]. They are designed to self-organize stable categories in response to on-line presentation of input. Built into these networks are nonlinear filters that perform noise quenching and feature enhancement [14,15]. The classifier architecture is logically divided into two fields: F_1, the bottom-up feature-representation field, and F_2, the top-down category-representation field. These fields should be differentiated from the layers of the multilayer perceptron because the individual fields of ART can be made up of several layers.

Each field has a logically distinct function. Nonlinear parallel interactions in the F_1 field suppress elements with low signal content and redistribute activity among the surviving elements,

thereby enhancing the contrast in the input signals. Bottom-up connections from F_1 prime the F_2 field for a category match. The first time a new input is experienced by the network, a parallel search of existing categories results; when the best match is found, a measurement analogous to the angle between the two vectors is compared to a single vigilance parameter. If this angle is close enough, the distance between the winner's direction and the input direction is decreased. If it is not close enough, the winner is disabled and the search reconvenes, terminating with either an existing category or, if no existing category is close enough, a new category. When the network becomes familiar with an input (typically a small number of exposures), the F_1-to-F_2 connections result in direct access to the representative category. The top-down connections from F_2 generate stable recognition codes for F_1 inputs. These codes are often referred to as category exemplars in the parlance of restricted-coulomb energy or RBF networks. For a detailed description of these networks, see Carpenter and Grossberg [14].

One advantage of ART networks over many other neural-network classifiers is that the ART networks do not have distinct training and testing modes; categories are constantly refined. Nevertheless, it is possible to guide ART networks in the early stages of operation by providing characteristic examples. This gives ART a head start in categorizing its data in much the same manner a teacher can help a student learn new information.

When taught in this way, another advantage of ART becomes apparent: it can take advantage of new training inputs immediately. In contrast, when one additional training example is added to networks such as the multilayer perceptron, the entire training set must be presented again, and training must be repeated from the beginning.

In order to efficiently teach and evaluate the performance in classifying seismic phases, ART (which produces unlabeled categories) was combined with a category labeler that generates a confusion matrix to represent the success of classification trials. This allows us to automate the evaluation process as a larger data base becomes available.

4.2.2 Gaussian Classification

A Gaussian classifier is relatively easy to implement using standard multilayer perceptrons [7], but the resulting software suffers from slow convergence and the retraining problems described above. A Gaussian classifier is still useful, however, since the results are well understood and it facilitates the design and testing of representations for waveform and parametric data.

The Gaussian classifier generalizes the concept of Euclidean distance by weighting the distance with the covariance matrix derived from the training set. The resulting metric (the Mahalonobis distance) can be used to determine the *a posteriori* probability of the signal identifications. Unfortunately, the Mahalonobis distance can be very sensitive to an outlying data point in the training set, particularly to mislabeled training items. These are both areas where neural-network classifiers such as ART may be able to help.

Using the principle component analysis cited earlier, parameter sets are generated which maximize the differences between pairs of classes. Although this readily allows a simple discriminant function (the Gaussian classifier), the parameter set transforms must be recomputed for each pairwise comparison. For example, one principal component analysis might generate a parameter vector well-suited for discriminating between *Pn* and *Pg* phases, but another analysis must be carried out to discriminate *Pn* phases from *Lg* phases. Similarly, identification among N possible phase classes requires $\frac{N!}{2(N-2)!}$ separate analyses and comparisons. However, this is manageable for $N \leq 6$, which is the case for phase identification.

In our implementation the final decision from among the pairwise decisions is made by a plurality vote whenever possible. For example, if *Pn* is the choice more often than any other phase, then *Pn* is selected. When there is no plurality winner, the decision is more complex and depends on the certainties of the binary decisions.

4.2.3 Radial Basis Functions

The RBF classifier [16] is a supervised neural- network-like multiclass classifier consisting of two stages. The first stage uses a set of radially symmetric "basis functions" to project the input vectors into a higher dimensional space, with the result that a classification problem cast into a high-dimensional space is more likely to be linearly separable than the same task in a lower dimensional space. The second stage operates from the high-dimensional space to perform the classification via discrimination hyperplanes. The orientation of the hyperplanes is set by supervised training, during which the misclassification errors are minimized for all elements of the training set simultaneously. The hyperplanes are represented by weights that can be learned incrementally in noisy environments, or that can be calculated analytically after a single pass of "clean" data using a matrix pseudoinverse to minimize the squared error [17]. The complexity of the classifier is controlled by the number of basis functions. As this number increases, a given data set is distributed more sparsely to classification space, and the classification hyperplanes fit the data more precisely (with fewer errors); at the same time, the classifier may overfit the data, in which case it is less successful with data not experienced previously. Thus we seek to use the least complex classifier which produces an acceptable misclassification error rate.

To maximize the usefulness of the basis functions, it is helpful to adjust their sensitivity to match the distribution of the inputs. If a particular input vector never occurs in a particular classifier application, it is wasteful to have part of the classifier be sensitive to it. The sensitivity of the basis functions is controlled by the locations of their centers (in input space) and their sizes (the coverage-extent in input space). A popular technique to determine the centers, which we also use, is to perform a cluster analysis using the k-means clustering algorithm [18], where k is the optimal number of basis functions determined empirically. The sizes of the basis functions can also be determined empirically, or can be based on the cluster standard deviations. Significant overlap among basis functions as well as extensive coverage of the input space are necessary for generalizability beyond the training data set.

4.3 Initial Experiments

4.3.1 Phase Identification Using Vertical Sonograms

Our initial experiments with sonograms examined several pattern-recognition preprocessing options, explored how much phase identification information might be contained in only the sonograms of vertical seismograms, and experimentally applied an ART2 approach to phase identification .

The results in this section are based on a subset of our initial waveform data set (See Section 2.1). The data consisted of single vertical seismograms from eight events (four earthquake and four explosions) recorded at NORSAR. Four phases (*Pn, Pg, Sn, Lg*) were reported for each of the events. For these experiments the phase onset times were picked by a Lincoln analyst. The waveform was cut into time blocks starting 1.6 seconds before each phase and continuing until the next phase, or 25 seconds later in the case of the *Lg* phase; noise and coda intervals were also used. A noise sample was taken from just before the *Pn* window and the coda was taken starting 45 seconds after the onset of the *Lg* phase. The 32 phases varied in duration from about 3.5 seconds to about 35 seconds.

The signals for each phase were subjected to the following processing. We first generated sonograms, which were then processed to enhance areas of rapid change, to emphasize the phase onset, to normalize the spectra by background noise, or to reduce sensitivity to source spectra by coda normalization. The processed images were then applied to an ART2 [14] network. A supervised labeler was used to map the ART2 output nodes into phase names. Later experiments will examine other neural-net classification techniques; we were initially interested in examining the effect of the sonogram-processing schemes on classification performance.

Altogether, we examined the 9 processing schemes listed in Table 3. They were applied in the order in which they are listed in the table from left to right. In all cases the images were smoothed and resampled along the frequency and time axes to avoid excessive sensitivity to slight shifts of the sonograms in frequency or time.

The sonogram images were given phase labels using an unsupervised ART2 classifier with a supervised phase labeler at its output. The number of elements in the ART2 input vectors was 240. The ART2 input vectors were obtained by raster scanning the processed sonogram images. Most of the sonograms contained more than 240 pixels and were down-sampled to get 240 values. One sonogram was smaller and was upsampled.

In our initial experiments we forced our ART2 classifier to generate fixed numbers of classes and compared the classification performance for the different processing schemes and for different numbers of ART2 categories. When the number of ART2 categories was greater than four (corresponding to the four phase types to be classified: *Pn, Pg, Sn, Lg*) the post-ART2 labeler was used to make the required many-to-one mapping. The number of ART2 categories was controlled by the "vigilance" parameter, which controls how alike inputs must be to be assigned to a single category. There is a lower limit to the number of categories that can be obtained by adjusting vigilance. In

TABLE 3

Sonogram Postprocessing Schemes

Name	Normalization		Data Compression			Image Processing Technique		
	Div. by Avg Noise	Div. by Avg Coda	Freq. Smooth & Sample	Time Smooth & Log. Sample	Log. of Power	DOG 1x1-3x3	DOG 3x3-5x5	Zero Crossing of DOG
ft			X	X				
ftl			X	X	X			
nft	X		X	X				
nftl	X		X	X	X			
cft		X	X	X				
cftl		X	X	X	X			
cftld		X	X	X	X	X		
cftlD		X	X	X	X		X	
cftle		X	X	X	X			X

X indicates computation performed; order is left to right.

ART2 there is an unassigned category which has arbitrary preassigned weights; it is always possible that it will be closer to a novel input than any of the previously assigned categories. Thus, if a classifier already has assigned categories and a new input is applied, a new category may be created regardless of the vigilance setting. The interpretation of this is that the data is sufficiently distinct to require additional categories to represent it.

Table 4 shows experimental results for the cases of four, eight, and 10 ART2 classes. Since we were concerned with four phase types, forcing ART2 to generate four categories seemed reasonable. However, in several cases it was not possible to produce only four categories by adjusting vigilance; four seemed too restrictive. Eight is another natural choice because our data included both earthquakes and explosions. The table also gives results for 10 ART2 categories; there is no need to restrict the number of ART2 categories to a number that we believe to be natural, and more than the natural number of categories may improve performance.

The results shown in Table 4 are significantly better than chance (25 percent). Many of the data representations force ART2 to create a minimum of 5 or 6 categories, because postprocessed data clusters are farther from each other than the unassigned category. Thus, ART2 tells us that we must use more than four categories to represent this data. Several processing paths cause

TABLE 4

Postprocessing Performance

Experiment Name	Results		
	4 Categories	8 Categories	10 Categories
ft	—	66%	66%
ftl	—	59%	63%
nft	—	59%	75%
nftl	47%	44%	50%
cft	—	78%	81%
cftl	—	84%	88%
cftld	69%	72%	78%
cftlD	—	78%	84%
cftle	56%	56%	72%

— indicates ART2 naturally produces more than 4 categories.

TABLE 5

Phase Confusion Matrix for cftl in 10 Categories

Desired Phase	Classified Phase			
	Pn	Pg	Sn	Lg
Pn	4	4	0	0
Pg	0	8	0	0
Sn	0	0	8	0
Lg	0	0	0	8

80 percent of the phases to be correctly classified; the best is 88 percent accurate. Allowing ART2 to use additional categories did improve the classification accuracy. We did not spend time tuning to this data base, because its small size may not generalize well to larger data bases. However, these early results are promising.

We next briefly examine which phases are confused with each other. Table 5 illustrates the misclassifications made for the example "cftl" and 10 ART2 categories. *Pn* and *Pg* are the only misclassified phases in this case. This is consistent with seismological expectations. Deciding between *Pn* and *Pg* on the basis of signal characteristics alone is more difficult than, for example, deciding between a *P*-type phase and either *Sn* or *Lg*.

Future work will build on these preliminary experiments. We will investigate sonogram-based ART2/hybrid classifier characteristics using a larger data base. The waveform data base being provided by SAIC (see Section 2.2) will be used. We will employ only automatic algorithms (no analyst intervention) to select phase onset times and locate coda segments. Another important task will be to explore the use the ART2/hybrid approach for phase classification using parameters (signal features) available within the IMS. We will develop networks and investigate classification performance using the IMS parameters alone and with sonograms.

4.3.2 Phase Identification Using Three-Component Autoregressive Models

Some of the results in this section are based on the initial waveform data set (built for HNC) and some are based on the high-quality analyst-corrected 50-waveform data base recently delivered to us by SAIC.

Using the multichannel autoregressive parameters, we applied two mechanisms to multichannel waveform data to perform phase identification: (1) Gaussian classification based on pairwise KL-transformed data and (2) RBF classification.

Preliminary experiments with 10 events and 40 phases from the initial SAIC/HNC waveform data base showed that the Gaussian classifier (and the ART classifier) performed at chance levels on the multichannel autoregressive parametric data. Although this was discouraging, the addition of the Karhunen-Loève (KL) transform (PCA) brought the classifier error rate down to less than 10 percent. There were four categories (*Pn*, *Pg*, *Sn*, and *Lg*). All 40 phases were used for training, so 10 percent is the error rate on the training set. Data-set limitations restricted the experiments to first-order three-component autoregressive models. Phase-onset times were provided by an analyst and the entire signal from the onset to the next phase onset (or 25 seconds, whichever was smaller) was used to estimate the autoregressive coefficients. The error rate is comparable to that obtained using ART2 and vertical sonograms.

Additional experiments have been performed using the 50-event waveform data base that was recently provided (see Section 2.2) by SAIC. Results obtained using 152 phase arrivals from the ARCESS array are presented in Table 6; it presents Gaussian classifier results for (1) when all the data were used for both training and testing and (2) when the 152 arrivals were split into separate

TABLE 6

Karhunen-Loève/Gaussian Classifier for Phase ID

% Correct			
Training*		**Test Data****	
Exact	**First-letter**	**Exact**	**First-letter**
51%	63%	insuff.	59%
* 152 phases used for training and testing			
** 100 phases for training and 52 for testing			

training and testing data. The automatic phase-arrival times provided by IMS were used as the start of the window in which autoregressive coefficients were calculated. A three-second window was used in all cases for estimating autoregressive coefficients.

There were six classes for these experiments: *Pn*, *Pg*, *Px*, *Lg*, *Sn*, and *Sx*. Phases labeled *Px* or *Sx* may have been too ambiguous to label more exactly, and so were treated as distinct labels during training. During recognition, however, we examined the case where the test-phase identification was required to match the analyst label exactly (i.e., $Pn = Pn$), and the case where only first letter matches were required (i.e., $Pn = Pg = Px$). As expected, success improved when the matching criterion was relaxed.

The number of KL coefficients (eigenvalues) used for classification ranged from 2 to 25, depending on which two phases were being distinguished. It was necessary to use fewer coefficients when one phase type of the pair was underrepresented in the data (e.g., there were only three phases labelled *Pg*). In the case when the data were split into separate testing and training sets, underrepresentation became so acute that there were insufficient data to design a classifier to distinguish all six phases; only "first letter" classification was possible. When there were enough data so that we could have generated more than 25 coefficients, we arbitrarily chose to use only 25. The performance impact of the underrepresented phase types in the data and of the explicit inclusion of the ambiguous *Sx* and *Px* types is uncertain and needs investigation.

For the RBF method, we used the first 25 parameters from the autoregressive model, giving a $d = 25$ dimensional input space. Since each parameter matrix is 3×3, this corresponds to about a third-order model. The number of phase identification classes was again six. The number of basis functions, f, was allowed to vary as we searched for an efficient representation. In Table 7, the complexity (i.e., the number of RBFs) increases in multiples of d (the input dimensionality).

TABLE 7

Radial Basis Functions for Phase ID

Complexity	% Correct			
RBF's	Training*		Test Data**	
f/d	Exact	First-letter	Exact	First-letter
1	***	***	21%	43%
2	67%	71%	31%	50%
3	76%	82%	15%	39%
4	89%	95%	17%	37%
5	97%	97.5%	29%	52%

* 152 phases used for training and testing

** 100 phases for training and 52 for testing

*** This case was not run.

These results show that the RBF classifier has greater success representing the training data as the classifier complexity increases, but the generality of the representation may suffer from overfitting the training data. In these initial experiments, the RBF easily outperformed the Karhunen-Loève transform(KLT)/Gaussian classifier during the training phase, but could not match the KLT/Gaussian classifier with the subsequent test data regardless of the classifier complexity.

We will need to improve on these success rates if we are to provide a useful capability for the IMS. For example, a quick check of the parametric data base maintained at the Center for Seismic Studies [5,6] suggests that only 20–30 percent of the phase labels assigned automatically by the IMS are subsequently corrected by the analyst. Based on this, a success rate less than 70–80 percent would not be of much value. We plan to review the CSS data base in more detail to fully understand the analyst correction statistics.

All of these results were obtained using a single three-component sensor that was not rotated to include radial and transverse components. For the later experiments only one three-second time window starting at the phase-onset time was used to generate autoregressive parameters. Alternative windows, sensor rotation, and multiple sensors are three options for improving performance. A larger or later window could provide a larger SNR or a view of the phase when it is in purer state. An azimuth estimate obtainable from three-component data [19,11] could be used to rotate

the sensors and develop an autoregressive representation that is more invariant to direction of arrival. Alternatively, the IMS frequency-wavenumber direction estimate could be used to provide the azimuth information needed to rotate the sensors. Data from more than one three-component sensor could also be combined, by beamforming or by correlation matrix averaging [11], to improve SNRs. We plan to explore these options and investigate the use of feature vectors that include polarization, speed, and other parameters that are automatically produced by the IMS. We expect to achieve substantial performance improvements by these means.

REFERENCES

1. T.C. Bache, et al., "Intelligent array system," Final Technical Report SAIC-90/1437, Science Applications International Corporation, San Diego, CA (October 1990).

2. T.C. Bache, S.R. Bratt, J. Wang, R.M. Fung, C. Kobryn, and J. Given, "The intelligent monitoring system," *Bull. Seis. Soc. Am.* **80**(6), Part B, 1833–1851 (1990).

3. S. Mykkeltveit and H. Bungum, "Processing of regional seismic events using data from small-aperture arrays," *Bull. Seis. Soc. Am.* **74**, 2313–2333 (1984).

4. S. Mykkeltveit, F. Ringdal, T. Kvaerna, and R.W. Alewine, "Application of regional arrays in seismic verification research," *Bull. Seis. Soc. Am.* **80**(6), Part B, 1777–1800 (1990).

5. J. Anderson, W.E. Farrell, K. Garcia, J. Given, and H. Swanger, "Center for Seismic Studies version 3 data base: schema reference manual," Center for Seismic Studies Technical Report C90-1 (September 1990).

6. J. Anderson and H. Swanger, "Center for Seismic Studies Version 3 data base: SQL tutorial," Center for Seismic Studies Technical Report C90-2 (September 1990).

7. R.P. Lippman, "Pattern classification using neural networks," *IEEE Commun. Mag.*, 47–64 (1989).

8. M.A.H. Hedlin, J.B. Minster, and J.A. Orcutt, "The time-frequency characteristics of quarry blasts and calibration explosions recorded in Kazakhstan, USSR," *Geophys. J. Int.* **99**, 109–121 (1989).

9. H.P. Harjes and M. Joswig, "Signal detection by pattern recognition methods," *The Vela Programs: A 25 Year Review of Basic Research*, edited by A. Kerr and D.L. Carlson, pp. 579–584 (1985).

10. M. Joswig, "Pattern recognition for earthquake detection," *Bull. Seis. Soc. Am.* **80**(1), 170–186 (1990).

11. A. Jurkevics, "Polarization analysis of three-component array data," *Bull. Seis. Soc. Am.* **78** 1725–1743 (1988).

12. K. Funkunaga and W.L.G. Koontz, "Application of the Karhunen-Loève expansion to feature selection and ordering," *IEEE Trans. Comput.* **19**(4), 311–318 (1970).

13. G.A. Carpenter and S. Grossberg, "The ART of adaptive pattern recognition by a self-organizing neural network," *IEEE Computer*, 77–88 (1988).

14. G.A. Carpenter and S. Grossberg, "ART 2: self organization of stable category recognition codes for analog input patterns," *Appl. Opt.* **26**(23), 4919–4930 (1987).

15 S. Grossberg, "Adaptive pattern classification and universal recoding II: feedback, expectation, olfaction, and illusions," *Biological Cybernetics* **23**, 187–202 (1976).

REFERENCES
(Continued)

16. D.S Broomhead and D. Lowe, "Multivariable functional interpolation and adaptive networks," *Complex Systems* **2**, 312–355 (1988).

17. S. Renals and R. Rohwer, "Phoneme classification experiments using radial basis functions," *Proceedings of the International Conference on Neural Networks I*, Washington, DC, 461–467 (1989).

18. R.O. Duda and P.E. Hart, *Pattern Classification and Scene Analysis*, New York, NY: John Wiley and Sons (1973).

19. R.G. Roberts, A. Christoffersson, and F. Cassidy, "Real-time event detection, phase identification, and source location estimation using single three-component seismic data," *Geophysical Journal* **97**, 471–480 (1989).

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (*Leave blank*)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	30 November 1990	Semiannual Tech Summary – 1 May – 30 November 1990

4. TITLE AND SUBTITLE

Artificial Neural Networks for
Seismic Data Interpretation

5. FUNDING NUMBERS

C — F19628-90-0002
PE — 61101E
PR — 570

6. AUTHOR(S)

Richard T. Lacoss, Robert K. Cunningham,
Susan R. Curtis, Michael C. Seibert

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Lincoln Laboratory, MIT
P.O. Box 73
Lexington, MA 02173-9108

8. PERFORMING ORGANIZATION REPORT NUMBER

S/A-ANN-12-30-90

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Air Force
PL/GPEH
Hanscom AFB, MA 01731-5000

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

ESD-TR-91-058

11. SUPPLEMENTARY NOTES

None

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (*Maximum 200 words*)

This is the first Semiannual Technical Summary report of the MIT Lincoln Laboratory Artificial Neural Networks for Seismic Data Interpretation project.

14. SUBJECT TERMS

neural networks
seismic discrimination
back-propagation
adaptive resonance theory
radio basis functions
phase identification

15. NUMBER OF PAGES
42

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	SAR